# The Evolution of Early-Life Effects on Social Behaviour – Why Should Social Adversity Carry Over to the Future?

Bram Kuijper[1,3] & Rufus A. Johnstone[2,4]

15th January 2019

**Abstract**

Numerous studies have shown that social adversity in early life can have long-lasting consequences for social behaviour in adulthood, consequences that may in turn be propagated to future generations. Given these intergenerational effects, it is puzzling why natural selection might favour such sensitivity to an individual's early social environment. To address this question, we model the evolution of social sensitivity in the development of helping behaviours, showing that natural selection indeed favours individuals whose tendency to help others is dependent on early-life social experience. In organisms with nonoverlapping generations, we find that natural selection can favour positive social feedbacks, in which individuals who received more help in early life are also more likely to help others in adulthood, while individuals who received no early-life help develop low tendencies to helping others later in life. This positive social sensitivity is favoured because of an intergenerational relatedness feedback: patches with many helpers tend to be more productive, leading to higher relatedness within the local group, which in turn favours higher levels of help in the next generation. In organisms with overlapping generations this positive feedback is less likely to occur, and who received more help may instead be less likely to help others (negative social feedback). We conclude that the early-life social influences can lead to strong between-individual differences in helping behaviour, which can take different forms dependent on the life history in question.

**Affiliations** 1) Environment and Sustainability Institute, University of Exeter Cornwall Campus, Penryn TR10 9FE, United Kingdom
2) Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom
3) a.l.w.kuijper@exeter.ac.uk
4) raj1003@hermes.ac.uk

**Keywords:** predictive adaptive response, developmental plasticity, inclusive fitness, maternal effect, parental effect, intergenerational effect

**Running title:** Early life effects on social behaviour

# 1 Introduction

In many taxa, the social environment experienced during early life gives rise to predictable between-individual differences in adult social behaviour [1–5]. In many rodents, individuals who have received limited parental care also provide less parental care themselves to their own offspring [6, 7]. By contrast, wild mongoose female who received more help from escorts in early life are less likely to help themselves as opposed to reproduction [8]. Both examples hint at a form of developmental plasticity, in which social cues early in life lead to irreversible developmental switching [3, 5, 9], thereby affecting the expression of prosocial behaviour later in life [7, 10, 11].

The long-term consequences of early-life social adversity, which may even spill over into the next generation, raise the question of why sensitivity to early social experiences has evolved at all: if anything, one might expect that offspring are selectively favoured to buffer the effects of early life adversity, so that the transmission of adverse social behaviours to future generations is precluded [12]. To understand why developmental plasticity is nonetheless favoured by natural selection, a growing body of theoretical work therefore suggests that early life effects may be an adaptive response to information about potential future environments (e.g., 13–21, but see [22]), suggesting that social adversity during early life is indicative of future social adversity, thus favouring the development of a less social phenotype. While this explanation is intuitive, a key shortcoming of these models is that they have been exclusively formulated with abiotic environments in mind, whereas in the context of social behaviours, the future is shaped by the actions of individuals themselves. Consequently, these models cannot explain why individuals who have experienced social adversity early in life are selectively favoured to go on and subsequently create a socially adverse environment for their own offspring [3, 23].

To understand how early-life social experiences can lead to the intergenerational transmission of socially benign or adverse conditions, we develop an evolutionary model of a developmentally plastic social trait. We focus on the evolution of a helping in a patch-structured population, in which individuals make an irreversible decision early in life to develop either as a nonreproductive helper or as a potential reproductive adult; an individual's strategy determines the probability with which it develops as a helper rather than a breeder (e.g., [24–27]). The exact number of helpers recruited to a patch varies, with the expected number proportional to the average helping tendency expressed by local individuals. In line with the majority of the theoretical literature, we assume that helping is assumed to increase the fecundity of the reproductives in the local group [28]. The helping tendency expressed by a newborn can then evolve to become dependent on the number of helpers in the local patch present at the time of birth, reflecting the results of empirical studies in which developmental plasticity is based on the current social structure of the local group (e.g., [4, 29, 30]). We then study whether social behaviours are indeed likely to become sensitive to social experience in early life, and if so, what form such developmental plasticity takes.

# 2 The Model

We consider a demographically explicit model of a sexually reproducing metapopulation, in which breeders are distributed over infinitely many demes (Wright's infinite island model [25, 31]). Each deme contains $n_\mathrm{b}$ adult breeders, who are assumed to reproduce as simultaneous hermaphrodites for the reason of tractability. In addition to breeders, demes can also contain $j$ nonreproductive helping individuals, thereby positively affecting the fecundity of their reproductive patch mates. Throughout the main text, we assume that generations are nonoverlapping, while results with overlapping generations are given in section S2.6 of the Online Supplement. We assume that individual demes vary in the number $0 \leq j \leq n_{\mathrm{h,max}}$ of helpers that have successfully been recruited (see the paragraph "Life cycle" below). To assess whether social experiences in early life affect later-life helping, we then ask whether the decision of newborns to become helpers evolves to be dependent (i.e., developmentally plastic helping, [32]) on the number of helpers currently present on the patch. For the sake of comparison, we also study the evolution of helping that is independent of the number . Below, we provide a verbal summary of the life cycle, while an extensive description is given in Section 1 of the Online Supplement. See Table S1 for an overview of the symbols used in the analytical description of the model below. Finally, we checked results using stochastic individual-based simulations in finite populations, which give very similar results (see Figure S3).

## 2.1 Life cycle

Consider a focal mutant adult breeder who lives on a patch with $n_\mathrm{b}-1$ other breeders and $j$ nonreproductive helpers. It randomly chooses a mate from among the $n_b$ breeders in the local patch and subsequently produces a large number $f_j$ offspring. Here, fecundity $f_j$ is an increasing function of total amount of help received from the $j$ helpers, which we assume to be equally distributed over the $n_\mathrm{b}$ breeders present in the patch. A juvenile born from the mutant focal breeder will forego on reproduction and develop as a helper with probability $h_j^\bullet$, where $^\bullet$ indicates the helping tendency expressed by the mutant mother (differing slightly from the average helping tendency $h_j$ in the population). As mentioned before, a key assumption of our analysis is that the tendency to develop as a helper can evolve to become dependent on the current number of helpers in the local patch $j$. Alternatively, with probability $1-h_j^\bullet$, a juvenile does not develop as a helper, in which case it either disperses to a randomly chosen remote patch with probability $d$ or remains at the local patch with probability $1-d$.

After juvenile dispersal, all non-helping juveniles, both philopatric and immigrant, then compete for the expected number of $n_\mathrm{b}$ vacant of breeding positions. The cycle then repeats, with the newly established breeders' fecundity now affected by a number of $k$ helpers, recruited from the helping juveniles born in the local patch. Specifically, we assume that the number of helpers in each patch is given by a truncated Poisson distribution, where the probability $s_{j\to k}(h_j)$ that a local patch which previously contained $j$ helpers now contains $k$ helpers

is given by

$$s_{j\to k}\left(h_j^{\circ}\right)=\begin{cases}\frac{\bar{n}\left(h_j^{\circ},j\right)^{k}\mathrm{e}^{-\bar{n}\left(h_j^{\circ},j\right)}}{k!} & k<n_{\mathrm{h,max}}\\ \sum_{\ell=n_{\mathrm{h,max}}}^{\infty}\frac{\bar{n}\left(h_j^{\circ},j\right)^{\ell}\mathrm{e}^{-\bar{n}\left(h_j^{\circ},j\right)}}{\ell!} & k=n_{\mathrm{h,max}}\end{cases}, \tag{1}$$

where the first line reflects the Poisson probability of recruiting $k$ helpers when $\bar{n}_h(h_j^{\circ},j)=n_{\mathrm{b}}f_j h_j^{\circ}$ is the average number of helping juveniles produced by all adult breeders in the local patch. The second line reflects the probability that the maximum of $k=n_{\mathrm{h,max}}$ helpers is attained, which occurs when $\ell=n_{\mathrm{h,max}}$ helpers are sampled, or when more helpers than positions available for them are sampled (i.e., $n_{\mathrm{h,max}}<\ell<\infty$), in which case we assume that helpers compete amongst themselves for the $n_{\mathrm{h,max}}$ available helping positions, with the unsuccessful helpers dying afterwards. After $k$ helpers have been recruited to the local patch, the cycle then repeats.

## 2.2 Fitness

The expected number $w_{ij}$ of offspring who successfully establish themselves in a patch with $i$ helpers and are born from a mutant adult breeder in a patch with a total number of $j$ helpers is then given by

$$w_{ij}=f_j\left(1-h_j^{\bullet}\right)\left[\frac{n_{\mathrm{b}}(1-d)\,s_{j\to i}\left(h_j^{\circ}\right)}{C\left(h_j^{\circ};\mathbf{h},j\right)}+d\sum_{k=0}^{n_{\mathrm{h,max}}}u_k\frac{n_{\mathrm{b}}s_{k\to i}(h_k)}{C(h_k;\mathbf{h},k)}\right], \tag{2}$$

where $f_j$ reflects the total number of surviving newborns produced by the focal adult breeder, a proportion $1-h_j^{\bullet}$ of which develop as juvenile reproductives (rather than helpers). These juvenile reproductives then go on to compete for any of the $n_{\mathrm{b}}$ available breeding positions in the natal patch with probability $1-d$ (first part in straight brackets), or in a random, remote patch with probability $d$ (second part in straight brackets), where $u_k$ reflects the population-wide frequency of patches currently containing $0\leq k\leq n_{\mathrm{h,max}}$ helpers. Philopatric reproductives compete with a total number of $C(h_j^{\circ};\mathbf{h},j)$ philopatric and immigrant offspring (see eq. [S2] in the Online Supplement), which is a function of (i) the average tendency $h_j^{\circ}$ expressed by any locally born newborn to develop as a helper, (ii) the population wide tendencies $\mathbf{h}=[h_0,h_1,\ldots,h_{n_{\mathrm{h,max}}}]$ to become helpers in any remote patch and (iii) the current number of helpers $j$ in the local patch. Finally, after successful establishment, the probability that the newly established breeder is accompanied by $i$ helpers in the next generation is then given by $s_{j\to i}(h_j^{\circ})$ (see eq. 1). The expected number of offspring who successfully compete in remote patch can then be derived in a similar fashion.

## 2.3 Evolutionary dynamics

We use a direct fitness method (also called neighbour-modulated fitness [33, 34]) to calculate evolutionary change $\mathcal{H}_k$ in the tendency to help when born on a patch containing $k$ helpers. According to a standard result [35–37], $\mathcal{H}_k$ is then given by

$$\mathcal{H}_k = V_k \sum_{i=0}^{n_{\mathrm{h,max}}} \sum_{j=0}^{n_{\mathrm{h,max}}} v_i u_j \left[ \frac{\partial b_{ij}}{\partial h_k^{\bullet}} + \frac{\partial b_{ij}}{\partial h_k^{\circ}} r_{\mathrm{local,j}} \right] \Bigg|_{\mathbf{h}^{\bullet}=\mathbf{h}^{\circ}=\mathbf{h}}, \tag{3}$$

where $V_k$ is a term that is proportional to the amount of additive genetic variance in the helping tendency $h_k$. Next, $v_i$ and $u_j$ are the individual reproductive values and stable class frequencies of adult breeders which are in a patch with $i$ helpers, which are obtained from the dominant left and right eigenvectors of the resident transition matrix (see eq. [S4]). The first partial derivative of the element $b_{ij}$ of the mutant transition matrix $\mathbf{B}$ (see eq. [S3]) reflects selection on offspring born from the focal mutant breeder, who express a helping tendency $h_k^{\bullet}$. The second partial derivative reflects selection on all breeders in the mutant's local patch, whose offspring, on average, express the helping tendency $h_k^{\circ}$. Finally, the relatedness coefficient $r_{\mathrm{local},j}$ reflects the relatedness between a focal adult breeder and any adult breeding individual (including the focal itself) in the local patch that currently contains $j$ helpers (see eq. [S7]). As we have not been able to find analytical solutions to find $h_k$, we developed an algorithm in C++ (source code available at `https://doi.org/10.5281/zenodo.1421729`) to numerically find the convergence stable values of the helping tendencies $\mathbf{h}$ (see Online Supplement). We also ran individual-based simulations to check results (see Figure S3), which showed that evolutionary branching of helping tendencies did not occur.

Throughout, we assume that helper-dependent fecundity of a focal breeder in a patch with $j$ helpers is given by the function $f_j = (1/n_{\mathrm{b}})(\phi_0 + \phi_1 j^{\phi_2})$, where $\phi_0$ is the baseline productivity of a patch without helpers, $\phi_1$ is the strength with which productivity increases with increasing helper number and $\phi_2$ reflects whether productivity increases in a linear, accelerating or decelerating fashion. We assume that the benefits of helping are equally shared among all $n_{\mathrm{b}}$ breeders.

# 3 Results

## 3.1 Result 1: Early-life social experiences strongly affect helping behaviour

To assess how the presence of helpers in early life affects an individual's tendency to help others, we focus on a scenario were maximally $n_{\mathrm{h,max}} = 5$ helpers can be recruited to a local patch and where helper number has a linear effect on local productivity. Results are, however, robust to different values of $n_{\mathrm{h,max}}$ (Figure S1) or cases where helper numbers increase local productivity in a decelerating fashion (Figure S2). We focus on $n_{\mathrm{b}} = 2$ breeders per patch;

higher values of $n_b$ result in lower values of local relatedness, so that helping evolves across a narrower range of parameter space. Yet when helping evolves, outcomes are qualitatively similar (results not shown).

Figure 1A shows that early-life effects on the development of adult helping behaviours are adaptive, as the probability of helping in adulthood is strongly dependent on the amount of help received in early life (as measured by the number of helpers in the local patch). Moreover, we find that those individuals who have experienced an intermediate number of helpers at birth (e.g., $n_h = 1, 2, 3$) are most likely to develop as helpers themselves in later life, whereas adults who have experienced either a very large amount of help (e.g., $n_h = 5$), and particularly those who have received no help at all ($n_h = 0$) are less likely to become helpers themselves. We also find that early-life effects extend the parameter space over which helping evolves (relative to populations which help unconditionally): for example, for high rates of dispersal ($d = 0.6$), only developmentally plastic helping evolves (compare solid vs dotted green lines in Figure 1A). Finally, because of the flexibility of developmentally plastic helping, we find that the population-wide average tendency to help is, in fact, lower for populations with developmental plasticity (see Figure S4). This is because developmentally plastic helpers express a low helping tendency in patches with no helpers, whereas unconditionally helping individuals express the same levels of helping everywhere (Figure 1A).

**Local variation in relatedness versus saturation drives plasticity in helping** To understand the evolution of developmental plasticity in helping behaviours, Figure 1B depicts the coefficient of consanguinity between two distinct breeders for patches with different numbers of helpers. We find that, once helping evolves, relatedness is highest in those patches which contain the largest numbers of helpers and lowest in patches where helpers are absent. This is because a larger number of helpers increases the fecundity of the local group, so that any vacant breeding spots are more likely to be claimed by locally born juveniles (rather than by remotely born offspring). Consequently, helpers and breeders are more likely to be related in the next generation, thus favouring stronger helping tendencies in patches that currently contain high numbers of helpers.

As relatedness is higher in patches with a large number of helpers, we would thus expect that helping is the highest in patches currently containing the largest numbers of helpers. At the same time, however, high productivities of patches with large numbers of helpers result in a rapid saturation of the available number of helper vacancies, explaining why helping tendencies are only maximized on patches with an intermediate number of helpers (see Figure 1A). We therefore conclude that local relatedness (favoring more helping) and saturation of helping positions (favoring less helping) are two opposing forces that determine the shape of the social reaction norm. Indeed, when saturation of helper positions is less important relative to local relatedness (for example, because of a higher total number of helper positions), developmentally plastic helping levels attain higher maximum levels and are maximized for patches with higher numbers of helpers (see Figure S1).

**Overlapping generations** When generations are overlapping, we find even stronger differences in helping tendencies between individuals who experienced few versus many helpers during early life (Figure S5B-D). However, in sharp contrast to populations with nonoverlapping generations (Figure 1A, Figure S5A), we find a negative relationship, so that individuals who received little help are more likely to help others, while those who received a lot of help are less likely to help others. The reason for this negative relationship between early life and later life help is that generational overlap changes the relative importance of local relatedness versus saturation of helping positions: while saturation of helping positions still favors reduced helping tendencies in patches with a higher number of helpers, the effect of local relatedness (which favors higher levels of help with increased helper number) is weakened. Local relatedness is less important because helpers will help their surviving parents (with whom relatedness is always high, regardless of the number of local helpers in the patch). Consequently, the effect of patch saturation prevails when generations overlap, so that individuals help most when having received little help themselves in early life.

## 3.2 Result 2: Helper presence predicts a social future

Our model also allows us to study the longer-term consequences that result from the presence (or absence) of helpers. When helper development depends on the current amount of help received (Figure 2), we find that even for relatively high levels of dispersal (i.e., $d \approx 0.6$) the current number of helpers is highly predictive of the number of helpers in the future (see also the autocorrelations in Figure 3, blue line). For example, patches that currently have no helpers ($n_{\mathrm{h}}(t) = 0$) are extremely unlikely to recruit any helpers in the future, while patches that currently have the maximum number of helpers (e.g., $n_{\mathrm{h}}(t) = 5$) are likely to have a large number of helpers again in the future.

By contrast, when helping is unconditional, help only evolves when dispersal is more limited (i.e., $d < 0.45$), but even then the number of helpers experienced in the current generation is a poorer predictor of the amount of help received in the future (see also Figure 3). Only in populations where dispersal becomes strongly limited, relatedness is high overall, so that helping evolves regardless of the current number of helpers in the local patch. Indeed, when dispersal becomes more strongly limited, the autocorrelations for developmentally plastic and unconditional helping become very similar.

When generations are overlapping, however, the presence of helpers in the current timestep is much less predictive of the number of helpers recruited in the future when compared to unconditional helping (see Figure S6). Autocorrelations are lower when generations are overlapping because of the aforementioned negative relationship between helper presence and help in later life (see Figure S5).

# 4 Discussion

Here, we have shown that the social environment experienced in early life can lead to substantial between-individual variation in adult socail behaviour. Moreover, we find that this developmental plasticity of social behaviour can take different forms dependent on the life-history in question: in taxa with nonoverlapping generations, we find that it can lead to positive social sensitivity where greater experience of prosocial behaviour in early life results in a greater tendency to behave prosocially in adulthood, while a reduction in prosocial behaviour is an adaptive response to social adversity in early life. By contrast, the opposite applies when generations are overlapping: here we find a negative social sensitivity, where prosocial behaviour in early life results in a lower tendency to behave prosocially in adulthood, while an *increase* in prosocial behaviour is an adaptive strategy to social adversity in early.

Our model predicts a positive relationship between early-life social experience and later-life social behaviour when generations are nonoverlapping, because helping promotes local productivity, thereby increasing local relatedness, and greater local relatedness in turn favours more helping. To understand this, focus on a local group with many helpers: this group will produce a large number of offspring, hence increasing the probability that a local breeding spot will be claimed by a locally born (rather than a remotely born) offspring in the next generation. In turn, this results in an increase in local relatedness (see Figure 1B), favouring a high tendency to develop as a helper. By contrast, patches which currently contain few helpers are less productive, ultimately resulting in a lower relatedness and a lower tendency to help. Of course, this kind of feedback between sociality and relatedness will only develop when individuals can adjust prosocial behaviour in response to juvenile cues that are predictive of local relatedness experienced as an adult. Our model shows that the experience of being helped in early life can serve as a reliable cue of expected relatedness in this way, thus driving developmental plasticity in later-life social behaviour.

Once positive social sensitivity has evolved, the intergenerational propagation of prosocial behaviour itself amplifies the benefits of helping, because an individual who becomes a helper not only boosts the fecundity of related breeders in the current generation, but also increases the tendency to help among progeny that remain on the local patch. Helping, in other words, ends up providing longer-term as well as shorter-term benefits. In a series of seminal models, Lehmann [38, 39] has previously shown that persistent benefits, which impact on the fitness of later generations, are particularly favourable for the evolution of helping, because they provide a partial escape from the constraints of local kin competition. These models, however, start from the assumption that the benefits of helping behaviour persist over time, as seems likely to be true for many beneficial modifications of the local environment such as construction or maintenance of a nest or burrow. Our model shows that even if helping has no such physically persistent effects, and only boosts the fecundity of breeders in the current generation, it may nevertheless end up yielding longer-term benefits because of the inter-generational propagation of prosocial tendencies.

In populations with overlapping generations, however, our model predicts a negative rela-

tionship between early-life social experience and later-life social behaviour. Although helping still promotes local productivity and local relatedness, local relatedness itself has a diminished effect on helping. This is because juvenile helpers are now more likely to help their surviving parents, rather than any members of their own generation, with whom relatedness increases with local helper numbers (see Figure 1B). Relatedness between a breeding parent and its offspring, however, is largely the same (0.5) unless inbreeding is very high. Consequently, help experienced during early life has little predictive value of parent-offspring relatedness that reflects the advantage of helping. However, the amount of help experienced during early life still predicts local productivity, where higher levels of local productivity imply that available helping positions become quickly satured. Consequently, due to diminished effect of local relatedness, early life help now favors a reduced tendency to help with increasing numbers of helpers, in order to avoid saturation of helping positions. Consequently, help received is indicative of both relatedness and local competition for helping positions, with the latter effect prevailing in taxa with overlapping generations, driving a negative feedback.

Previous models of the evolution of early life effects have focused chiefly on adaptation to fluctuations in the abiotic environment (e.g., [13, 15, 18, 40–43]) with surprisingly little attention given to social sensitivity (as previously noted in [3]). A key prediction of existing theory is that environmental conditions need to be sufficiently autocorrelated with later-life environmental conditions. However, some studies suggest that autocorrelations from climatic timeseries are, in fact, small and thus cannot readily account for the widespread occurrence of early-life effects (e.g., [22, 44]). Our model, however, shows that variation in the social environment can drive the evolution of early-life effects, even in the absence of autocorrelations in the abiotic environment, because social sensitivity itself generates high autocorrelations between parental and offspring social environments (see Figure 3). Hence, our study suggests that the social environment may in general play a more important role in the evolution of early-life effects than does the abiotic environment (see also [45, 46]).

Our model suggests a number of possible directions for future work: as discussed above, one key prediction is that increased levels of social behaviour result in increases in relatedness, thus creating a positive feedback loop (unless checked by increased saturation of helping positions). A typical consequence of positive feedback loops is that they often result in alternatively stable states [47]. Indeed, Figure 2 suggests that developmental plasticity may well result in a social polymorphism at the patch level, in which some populations may become locked into persistent low prosocial states, whereas others become locked into highly prosocial states. For example, when $n_h = 0$, the expected number of helpers in the next generation will be again 0 for $d = 0.4$ or $d = 0.6$ (see bottom left corner of Figure 2), so that the lack of help engenders little help in future. By contrast, when helpers are present in the local patch ($n_h \geq 1$), more helpers are likely to be recruited in the next generation (except when the current number of helpers is already at its maximum), engendering more help in future (Figure 2, rightwards). While neither state will persist indefinitely (due to demographic stochasticity in recruitment of helpers to a patch), positive transgenerational feedback on social behaviour will tend to maintain these differences between patches for longer than would otherwise be

the case. Such persistent polymorphism represents a group-level analog of the persistent differences in individual behaviour that emerge during early life and are maintained in models of personality evolution [48, 49]; one might even speak of the emergence of 'collective personalities'. However, a complete analysis of these consequences of developmental plasticity in helping is beyond the scope of the current paper and would merit further study.

Next, the current model focused on the evolution of helping behaviours, as this provides a straighforward context for modeling the evolutionary consequences of social adversity. However, it remains to be seen whether these conclusions generalize to other life history traits as well. In the context of parental care, for example, offspring born from mothers who provide high levels of care are more likely to survive, hence increasing relatedness in the local patch for similar reasons as in the current model. If mating is local, this results in an increase in parent-offspring relatedness, potentially favoring higher levels of parental care. By contrast, a parent who actively harms its offspring (e.g., maternal abuse in macaques [50]) may produce fewer surviving offspring, thus resulting in lower values of parent-offspring relatedness, which in turn may further enhance the evolution of harming. Formal models are thefore necessary to study the role of developmental plasticity in the face of other life history traits.

Finally, for reasons of tractability, the current study only focused on social traits characterized by plasticity that is irreversible, so that once an individual has committed itself to helping it will do so for the rest of its life. While such social traits may be particularly relevant for early forms of cooperative breeding in insects, many vertebrate societies are characterized by forms of reversible plasticity, where helpers become reproductives later in life [3], or reproductives become helpers (as in the case of grandparental care [51]). However, we would predict that the qualitative nature of our predictions may well be robust to the presence of reversible plasticity, in which case the numbers of helpers present in early life influence the total amount of time during an offspring's life that is devoted to help. Overall, our model shows that early-life effects in social contexts can be adaptive, but highlights the need for further study to understand their ecological significance.

**Acknowledgements** BK is funded by a Leverhulme Trust Early Career Research Fellowship (ECF 2015-273). We thank two reviewers for their comments on the manuscript. BK thanks Ana Duarte for discussions of early-life effects. We thank the University of Exeter for providing computational resources through the Carson computing cluster.

# References


[1] Kasumovic, M. M. & Brooks, R. C., 2011 It's all who you know: the evolution of socially cued anticipatory plasticity as a mating strategy. *Q. Rev. Biol.* **86**, 181–197. doi:10.1086/661119.

[2] Taborsky, B. & Oliveira, R. F., 2012 Social competence: an evolutionary approach. *Trends Ecol. Evol.* **27**, 679 – 688. doi:10.1016/j.tree.2012.09.003.

[3] English, S., Browning, L. E. & Raihani, N. J., 2015 Developmental plasticity and social specialization in cooperative societies. *Anim. Behav.* **106**, 37 – 42. doi:10.1016/j.anbehav.2015.05.006.

[4] French, J. A. & Carp, S. B., 2016 Early-life social adversity and developmental processes in nonhuman primates. *Curr. Opin. Behav. Sci.* **7**, 40 – 46. doi:https://doi.org/10.1016/j.cobeha.2015.11.004. Development and behavior.

[5] Taborsky, B., 2017 Developmental plasticity: preparing for life in a complex world. *Adv. Stud. Behav.* **49**, 49 – 99. doi:10.1016/bs.asb.2016.12.002.

[6] Curley, J., Davidson, S., Bateson, P. & Champagne, F., 2009 Social enrichment during postnatal development induces transgenerational effects on emotional and reproductive behavior in mice. *Front. Behav. Neurosci.* **3**. doi:10.3389/neuro.08.025.2009.

[7] Perkeybile, A. M. & Bales, K. L., 2017 Intergenerational transmission of sociality: the role of parents in shaping social behavior in monogamous and non-monogamous species. *J. Exp. Biol.* **220**, 114–123. doi:10.1242/jeb.142182.

[8] Vitikainen, E. I. K., Thompson, F. J., Marshall, H. H. & Cant, M. A., 2019 Live long and prosper: durable benefits of early life care in banded mongooses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **in press**.

[9] Nettle, D. & Bateson, M., 2015 Adaptive developmental plasticity: what is it, how can we recognize it and when can it evolve? *Proc. R. Soc. Lond. B Biol. Sci.* **282**, 20151 005. doi:10.1098/rspb.2015.1005.

[10] Blazei, R. W., Iacono, W. G. & Krueger, R. F., 2006 Intergenerational transmission of antisocial behavior: How do kids become antisocial adults? *Appl. Prev. Psych.* **11**, 230–253. doi:10.1016/j.appsy.2006.07.001.

[11] Champagne, F. A., 2016 Epigenetic legacy of parental experiences: dynamic and interactive pathways to inheritance. *Dev. Psychopathol.* **28**, 1219–1228. doi:10.1017/S0954579416000808.

[12] Wells, J. C. K. & Johnstone, R. A., 2017 Modeling developmental plasticity in human growth: buffering the past or predicting the future? In *The Arc of Life: Evolution and Health Across the Life Course* (eds. G. Jasienska, D. S. Sherry & D. J. Holmes). New York, NY: Springer, pp. 21–39. doi:10.1007/978-1-4939-4038-7_3.

[13] Frankenhuis, W. E. & Panchanathan, K., 2011 Balancing sampling and specialization: an adaptationist model of incremental development. *Proc. R. Soc. Lond. B Biol. Sci.* **278**, 3558–3565. doi:10.1098/rspb.2011.0055.

[14] Shea, N., Pen, I. & Uller, T., 2011 Three epigenetic information channels and their different roles in evolution. *J. Evol. Biol.* **24**, 1178–1187. doi:10.1111/j.1420-9101.2011.02235.x.

[15] Nettle, D., Frankenhuis, W. E. & Rickard, I. J., 2013 The evolution of predictive adaptive responses in human life history. *Proc. R. Soc. Lond. B Biol. Sci.* **280**. doi:10.1098/rspb.2013.1343.

[16] Burgess, S. C. & Marshall, D. J., 2014 Adaptive parental effects: the importance of estimating environmental predictability and offspring fitness appropriately. *Oikos* **123**, 769–776. doi:10.1111/oik.01235.

[17] English, S., Pen, I., Shea, N. & Uller, T., 2015 The information value of non-genetic inheritance in plants and animals. *PLoS ONE* **10**, e0116 996. doi:10.1371/journal.pone.0116996.

[18] Stamps, J. A. & Krishnan, V. V., 2014 Combining information from ancestors and personal experiences to predict individual differences in developmental trajectories. *Am. Nat.* **184**, 647–657. doi:10.1086/678116.

[19] Fawcett, T. W. & Frankenhuis, W. E., 2015 Adaptive explanations for sensitive windows in development. *Front. Zool.* **12**, 1–14. doi:10.1186/1742-9994-12-S1-S3.

[20] Kuijper, B. & Johnstone, R. A., 2016 Parental effects and the evolution of phenotypic memory. *J. Evol. Biol.* **29**, 265–276. doi:10.1111/jeb.12778.

[21] Frankenhuis, W. E., Panchanathan, K. & Belsky, J., 2016 A mathematical model of the evolution of individual differences in developmental plasticity arising through parental bet-hedging. *Dev. Sci.* **19**, 251–274. doi:10.1111/desc.12309.

[22] Wells, J. C. K., 2007 Flaws in the theory of predictive adaptive responses. *Trends Endocrinol. Metab.* **18**, 331–337. doi:10.1016/j.tem.2007.07.006.

[23] Taborsky, B., 2016 Opening the black box of developmental experiments: behavioural mechanisms underlying long-term effects of early social experience. *Ethology* **122**, 267–283. doi:10.1111/eth.12473.

[24] Lehmann, L., Ravigné, V. & Keller, L., 2008 Population viscosity can promote the evolution of altruistic sterile helpers and eusociality. *Proc. R. Soc. Lond. B Biol. Sci.* **275**, 1887–1895. doi:10.1098/rspb.2008.0276.

[25] Lehmann, L. & Rousset, F., 2010 How life history and demography promote or inhibit the evolution of helping behaviours. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 2599–2617. doi:10.1098/rstb.2010.0138.

[26] Gardner, A., West, S. A. & Wild, G., 2011 The genetical theory of kin selection. *J. Evol. Biol.* **24**, 1020–1043. doi:10.1111/j.1420-9101.2011.02236.x.

[27] Johnstone, R. A., Cant, M. A. & Field, J., 2012 Sex-biased dispersal, haplodiploidy and the evolution of helping in social insects. *Proc. R. Soc. Lond. B Biol. Sci.* **279**, 787–793. doi:10.1098/rspb.2011.1257.

[28] Rodrigues, A. M. M. & Kokko, H., 2016 Models of social evolution: can we do better to predict 'who helps whom to achieve what'? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150088. doi:10.1098/rstb.2015.0088.

[29] Russell, A. F. & Lummaa, V., 2009 Maternal effects in cooperative breeders: from hymenopterans to humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 1143–1167. doi:10.1098/rstb.2008.0298.

[30] Fischer, S., Bessert-Nettelbeck, M., Kotrschal, A. & Taborsky, B., 2015 Rearing-group size determines social competence and brain structure in a cooperatively breeding cichlid. *Am. Nat.* **186**, 123–140. doi:10.1086/681636.

[31] Wright, S., 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159.

[32] Holman, L., 2014 Conditional helping and evolutionary transitions to eusociality and cooperative breeding. *Behav. Ecol.* **25**, 1173–1182. doi:10.1093/beheco/aru100.

[33] Hamilton, W. D., 1964 The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16. doi:10.1016/0022-5193(64)90038-4.

[34] Taylor, P. D., Wild, G. & Gardner, A., 2007 Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* **20**, 301–309. doi:10.1111/j.1420-9101.2006.01196.x.

[35] Taylor, P. D., 1990 Allele-frequency change in a class-structured population. *Am. Nat.* **135**, 95–106. doi:10.1086/285034.

[36] Taylor, P. D. & Frank, S. A., 1996 How to make a kin selection model. *J. Theor. Biol.* **180**, 27–37. doi:10.1006/jtbi.1996.0075.

[37] Otto, S. P. & Day, T., 2007 *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton: Princeton University Press.

[38] Lehmann, L., 2007 The evolution of trans-generational altruism: kin selection meets niche construction. *J. Evol. Biol.* **20**, 181–189. doi:10.1111/j.1420-9101.2006.01202.x.

[39] Lehmann, L., 2008 The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. *Evolution* **62**, 549–566.

[40] Kuijper, B. & Johnstone, R. A., 2013 How should mothers adjust the size of their offspring to local environmental cues? *J. Evol. Biol.* **26**, 1488–1498. doi:10.1111/jeb.12156.

[41] Fischer, B., Van Doorn, G. S., Dieckmann, U. & Taborsky, B., 2014 The evolution of age-dependent plasticity. *Am. Nat.* **183**, 108–125. doi:10.1086/674008.

[42] English, S., Fawcett, T. W., Higginson, A. D., Trimmer, P. C. & Uller, T., 2016 Adaptive use of information during growth can explain long-term effects of early life experiences. *Am. Nat.* **187**, 620–632. doi:10.1086/685644.

[43] Kuijper, B. & Johnstone, R. A., 2018 Maternal effects and parent-offspring conflict. *Evolution* **72**, 220–233. doi:10.1111/evo.13403.

[44] Baig, U., Belsare, P., Watve, M. & Jog, M., 2011 Can thrifty gene(s) or predictive fetal programming for thriftiness lead to obesity? *J. Obesity* , 1–11doi:10.1155/2011/861049.

[45] Nettle, D., Bateson, M. & Cooper, R., 2017 Childhood and adult socioeconomic position interact to predict health in mid life in a cohort of british women. *PeerJ* **5**, e3528.

[46] Frankenhuis, W., Nettle, D. & Dall, S. R. X., 2019 A case for environmental statistics of early life effects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **in press**.

[47] Lehtonen, J. & Kokko, H., 2011 Positive feedback and alternative stable states in inbreeding, cooperation, sex roles and other evolutionary processes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 211–221. doi:10.1098/rstb.2011.0177.

[48] Wolf, M., van Doorn, G. S., Leimar, O. & Weissing, F. J., 2007 Life-history trade-offs favour the evolution of animal personalities. *Nature* **447**, 581–584. doi:10.1038/nature05835.

[49] Wolf, M., Van Doorn, G. S. & Weissing, F. J., 2008 Evolutionary emergence of responsive and unresponsive personalities. *Proc. Natl. Acad. Sci. USA*. doi:10.1073/pnas.0805473105.

[50] Maestripieri, D., 2005 Early experience affects the intergenerational transmission of infant abuse in rhesus monkeys. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9726–9729. doi:10.1073/pnas.0504122102.

[51] Croft, D. P., Brent, L. J., Franks, D. W. & Cant, M. A., 2015 The evolution of prolonged life after reproduction. *Trends in Ecology & Evolution* **30**, 407–416. doi:10.1016/j.tree.2015.04.011.

[52] Taylor, P. D., 1996 The selection differential in quantitative genetics and ESS models. *Evolution* **50**, 2106–2110. doi:10.2307/2410769.

[53] Dieckmann, U. & Law, R., 1996 The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612. doi:10.1007/BF02409751.

[54] Geritz, S. A. H., Kisdi, É., Meszéna, G. & Metz, J. A. J., 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57. doi:10.1023/a:1006554906681.

[55] Christiansen, F. B., 1991 On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* **138**, 37–50. doi:10.1086/285203.

[56] Kimura, M. & Crow, J. F., 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

# 5 Figure captions

**Figure 1** The evolution of developmentally plastic and unconditional helping behaviours in patches that contain $0 \leq n_{\mathrm{h}} \leq 5$ helpers in early life, for three different values of juvenile dispersal $d$ (panel A). When helping is developmentally plastic, individuals develop higher levels of help in patches where helpers are present in early life ($n_{\mathrm{h}} > 0$) relative to patches where help in early life is absent ($n_{\mathrm{h}} = 0$). Panel B depicts the corresponding relatedness coefficients, showing that relatedness is higher in patches with more helpers, as these are more productive (hence making it more likely that philopatric offspring claim breeding spots). Note that when $d = 0.6$, unconditional helping does not evolve, hence we only have $n_{\mathrm{h}} = 0$. The corresponding line in panel A is hence only drawn for the purpose of illustration, while we do not depict a corresponding relatedness coefficient for $d = 0.6$ and unconditional helping. See also Figure S1 for $n_{\mathrm{h,max}} \in \{3, 10, 20\}$ and Figure S2 when productivity increases in a decelerating (rather than linear) fashion with increasing numbers of helpers. Parameters: $n_{\mathrm{b}} = 2$, $\phi_0 = 1$, $\phi_1 = 5$, $\phi_2 = 1$, $n_{\mathrm{h,max}} = 5$.

**Figure 2** The expected number of helpers recruited in the next timestep increases with current number of helpers in the local patch. It does so much more rapidly in case helping is developmentally plastic (solid lines). Note that when $d = 0.6$, unconditional helping does not evolve, hence the expected number of helpers is always equal to 0. Parameters as in Figure 1.

**Figure 3** Temporal autocorrelation in patterns of presence versus absence of helpers between parental and offspring generations, for populations in which help is developmentally plastic (solid blue line) versus unconditional (dotted pink lines), while varying the probability of juvenile dispersal $d$. Parameters as in Figure 1.

Figure 1:

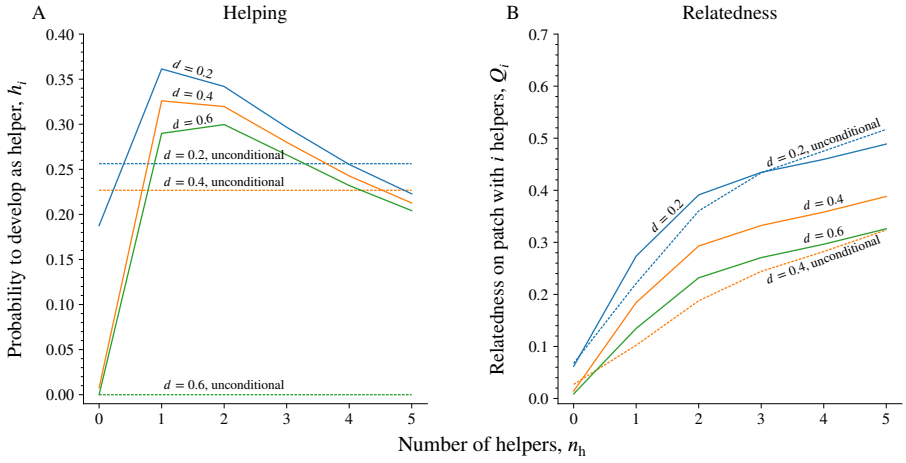


Figure 1:

Figure 2:

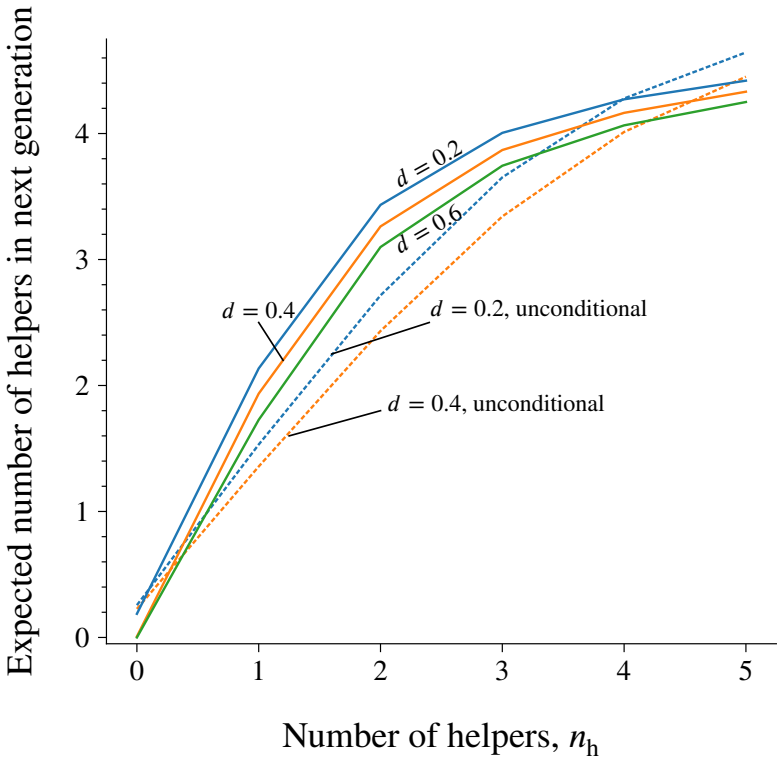


Figure 2:

Figure 3:

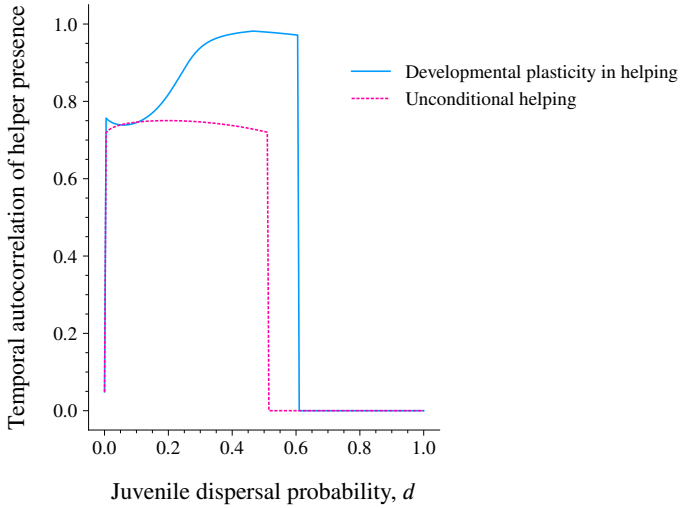


Figure 3:

# The Evolution of Early-Life Effects on Social Behaviour – Why Should Social Adversity Carry Over to the Future?

## Online Supplement

Bram Kuijper & Rufus A. Johnstone

## S1 Supplementary Figures

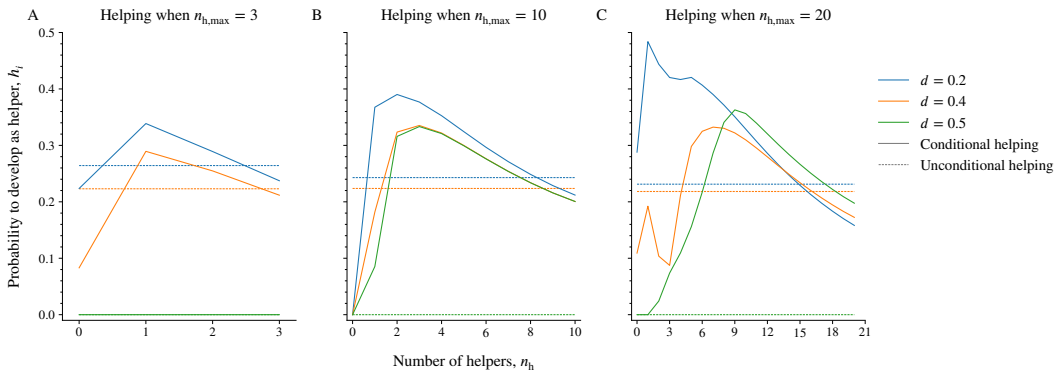


Figure S1:

**Figure S1** The evolution of the tendency to help $h_i$ while varying the maximum number $n_{h,max}$ of helpers that can be recruited to any local patch. Parameters: $n_b = 2$, $\phi_0 = 1.0$, $\phi_1 = 5$, $\phi_2 = 1.0$.

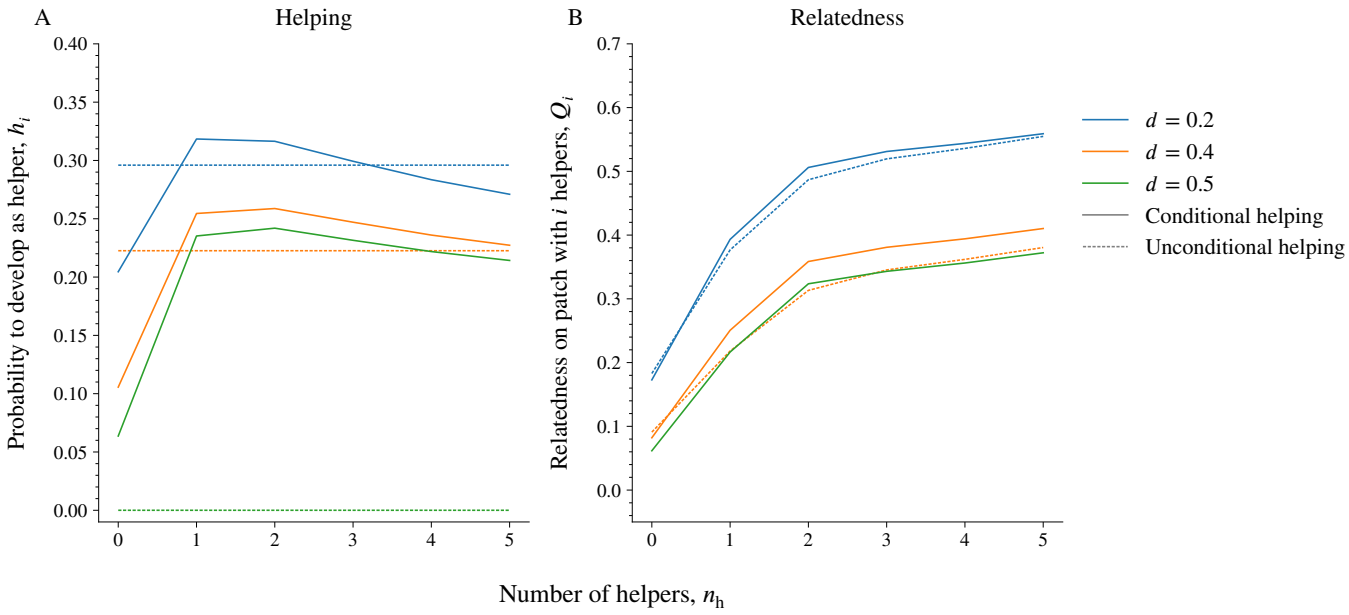


Figure S2:

**Figure S2** The evolution of the tendency to help $h_i$ when the increase in patch productivity decelerates with an increasing number of helpers ($\phi_2 = 0.5$). Results are qualitatively similar to Figure 1 in the main text, although developmental plasticity in helping is more modest. Note that we do not display relatedness for unconditional helping when $d = 0.5$ as helping does not evolve in this example. Parameters: $n_b = 2, n_{h,max} = 5$, $\phi_0 = 1.0$, $\phi_1 = 5.0$.

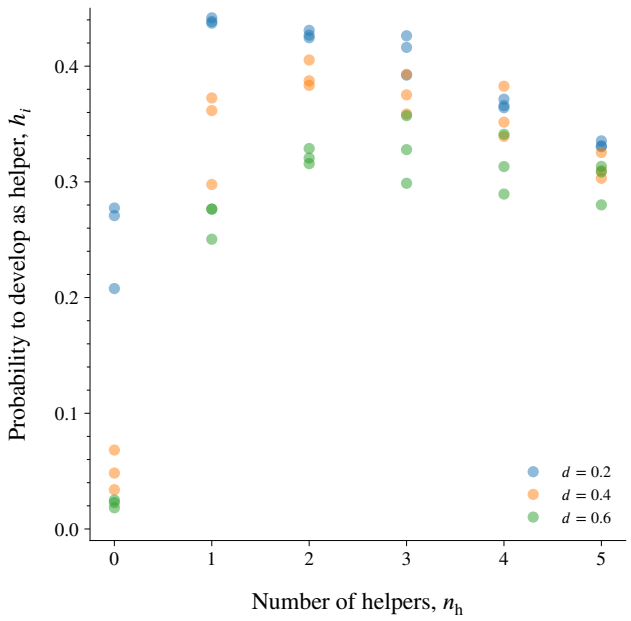


Figure S3:

**Figure S3** Stochastic, individual-based simulations show very similar results for the evolution of the tendency to help $h_i$ when compared to Figure 1 in the main text. Each dot depicts the population average value of $h_i$ evolved during a single simulation. Parameters: $n_b = 2$, $n_{h,max} = 5$, $\phi_0 = 0.4$, $\phi_1 = 5.0$, $\phi_2 = 1.0$. S

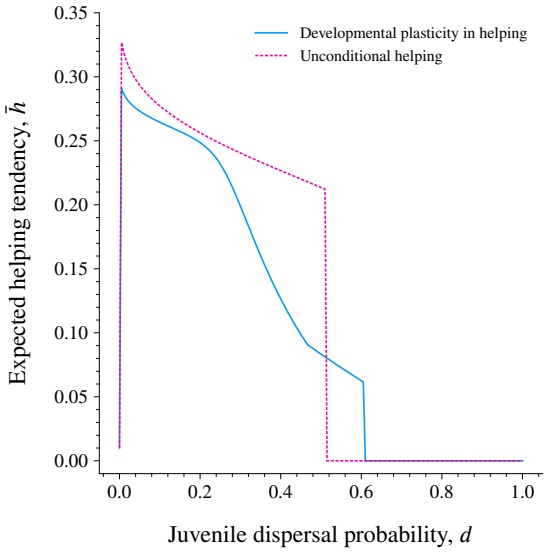


Figure S4:

**Figure S4** The average helping tendency $\bar{h}$ expressed by all mothers across the population for different levels of juvenile dispersal $d$. We find that unconditionally helping populations express higher average values of help, because all patches receive the same level of unconditional help. By contrast, patches with $n_{\mathrm{h}} = 0$ helpers receive little to no help in populations with developmental plasticity in helping (see Figure 1A in the main text). The average helping tendency was calculated as $\bar{h} = \sum_{i=0}^{n_{\mathrm{h,max}}} u_i h_i$, where $u_i$ is the frequency of a patch with $i$ helpers (see Table S1). Parameters as in Figure 1 in the main text.

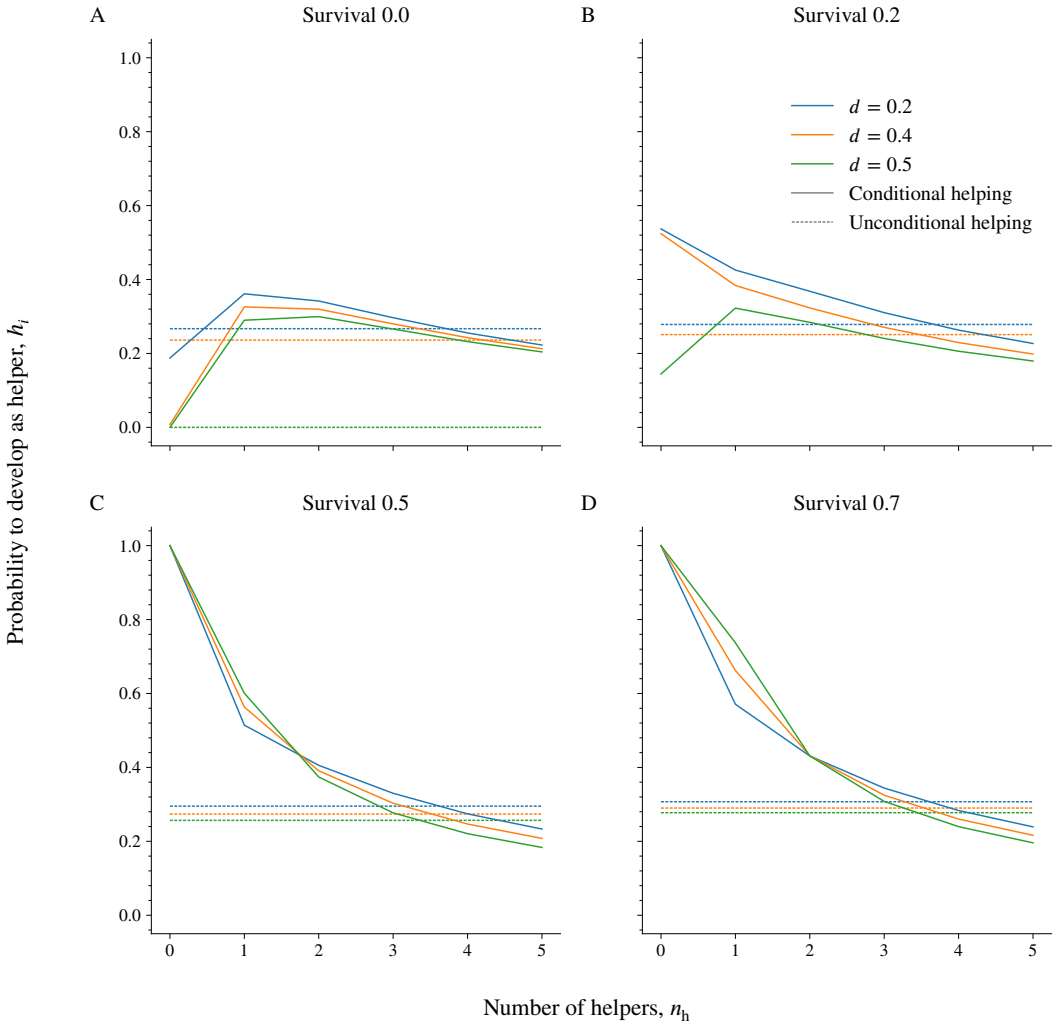


Figure S5:

**Figure S5** Effect of increasing amounts of generational overlap (measured by increased adult survival) on the evolution of developmental plasticity in helping tendencies $h_i$. In contrast to scenarios with nonoverlapping generations (panel A), overlapping generations cause helping tendencies to be very high in patches with no helpers, while helping tendencies are lowest in patches with many helpers. Parameters as in Figure 1. Panel A is Figure 1A of the main text and is depicted here for the sake of comparison.

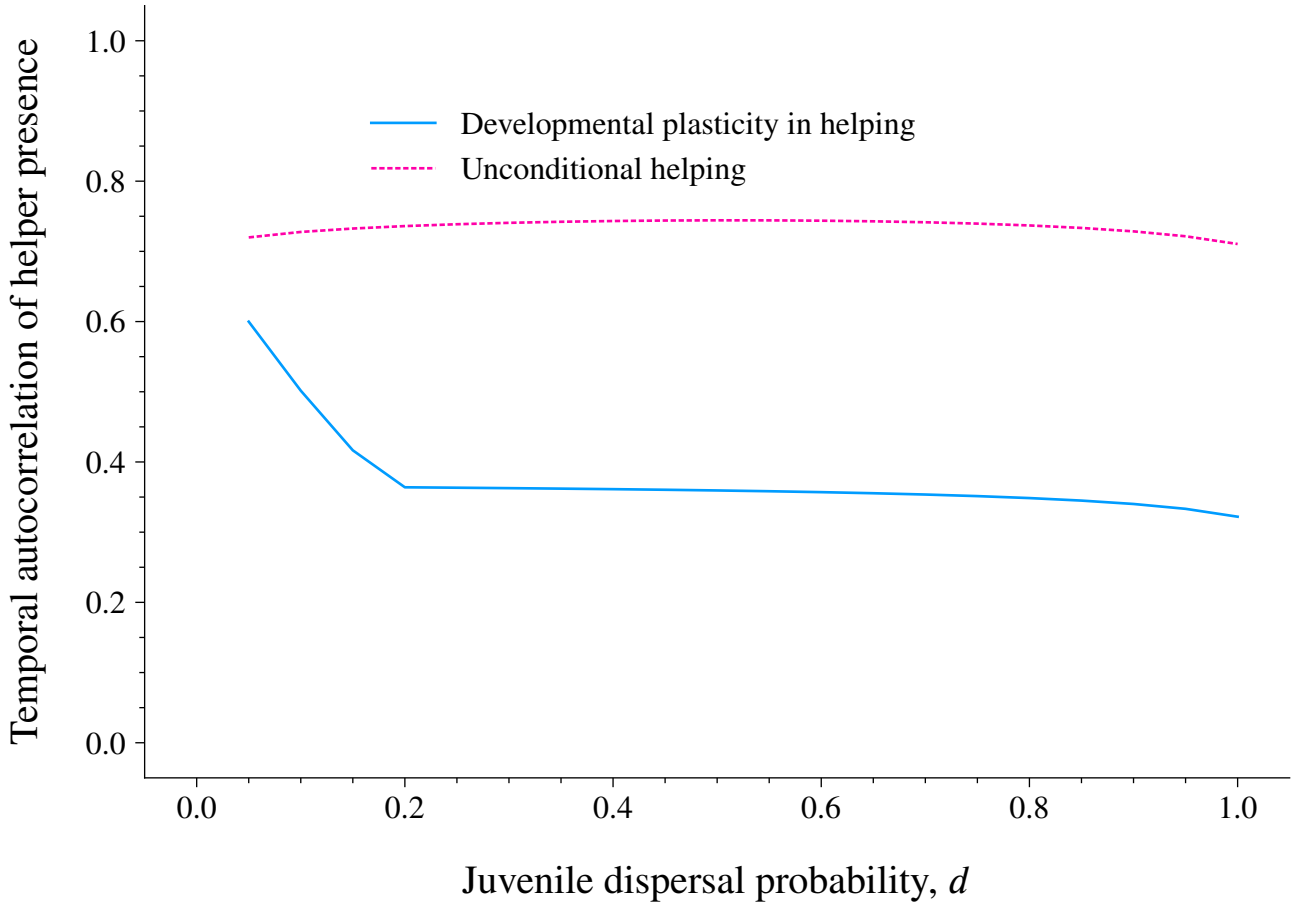


Figure S6:

**Figure S6** Autocorrelation in helper presence vs absence between parental and offspring generations, when generations overlap (adult survival during each timestep is $1-m=0.5$). In contrast to the scenario where generations are nonoverlapping (see Figure 3), helper presence/absence is a much poorer predictor of helper presence/absence in the future. This is because patches in which there is little help at time $t$ may recruit more helpers in time $t+1$ when generations are overlapping (see Figure S5C for $d=0.2$). Vice versa, patches in which there are already a lot of helpers may be less effective at recruiting more helpers in the future.

# S2 Description of the analytical model

Here we provide a more complete overview of the analytical model, where Table S1 provides an overview of the notation used.

## S2.1 Fitness expressions

To recap from the main text, the expected number $w_{ij}$ of offspring who successfully establish themselves in a patch with $i$ helpers and are born from a mutant adult breeder in a patch with a total number of $j$ helpers is given by

$$w_{ij} = f_j\left(1-h_j^{\bullet}\right)\left[\frac{n_{\mathrm{b}}(1-d)\,s_{j\to i}\left(h_j^{\circ}\right)}{C\left(h_j^{\circ};\mathbf{h},j\right)}+d\sum_{k=0}^{n_{\mathrm{h,max}}}u_k\frac{n_{\mathrm{b}}s_{k\to i}(h_k)}{C(h_k;\mathbf{h},k)}\right], \tag{S1}$$

where $f_j$ reflects the total number of surviving newborns produced by the focal adult breeder, a proportion $1-h_j^{\bullet}$ of which develop as juvenile reproductives (rather than helpers). These juvenile reproductives then go on to compete for any of the $n_{\mathrm{b}}$ available breeding positions in the natal patch with probability $1-d$ (first part in straight brackets), or in a random, remote patch with probability $d$ (second part in straight brackets), where $u_k$ reflects the population-wide frequency of patches currently containing $0 \leq k \leq n_{\mathrm{h,max}}$ helpers. Philopatric reproductives compete with a total number of $C(h_j^{\circ};\mathbf{h},j)$ philopatric and immigrant offspring (see eq. [S2] in the Online Supplement), which is a function of (i) the average tendency $h_j^{\circ}$ expressed by any locally born newborn to develop as a helper, (ii) the population wide tendencies $\mathbf{h}=[h_0,h_1,\ldots,h_{n_{\mathrm{h,max}}}]$ to become helpers in any remote patch and (iii) the current number of helpers $j$ in the local patch. Finally, after successful establishment, the probability that the newly established breeder is accompanied by $i$ helpers in the next generation is then given by $s_{j\to i}(h_j^{\circ})$ (see eq. 1). The expected number of offspring who successfully compete in remote patch can then be derived in a similar fashion.

The total number of competing reproductive juveniles in a local patch $C(h_j^{\circ};\mathbf{h},j)$ currently containing $j$ helpers (where $h_j^{\circ}$ reflects the average tendency to develop as a helper) is given by

$$C\left(h_j^{\circ};\mathbf{h},j\right) = n_{\mathrm{b}}f_j\left(1-h_j^{\circ}\right)(1-d)+d\sum_{k=0}^{n_{\mathrm{h,max}}}u_k n_{\mathrm{b}}f_k(1-h_k), \tag{S2}$$

where the first part of the expression above describes the total number of philopatric, non-helping offspring produced by all members of the local group. The second part is the total number of juveniles who migrate to the current patch (with probability $d$), originating from a patch that currently contains $k$ helpers (with probability $p_k$) resulting in a patch-level production of $n_{\mathrm{b}}f_k(1-h_k)$ reproductives.

The mutant transition matrix $\mathbf{B}$, with elements $b_{ij}=w_{ij}$, then comprises the transition

| Symbol | Explanation |
|---|---|
| $\mathbf{A}$ | Resident transition matrix |
| $a_{ij}$ | Entry in row $i$ and column $j$ of the resident transition matrix |
| $\mathbf{B}$ | Mutant transition matrix |
| $b_{ij}$ | Entry in row $i$ and column $j$ of the mutant transition matrix |
| $C(h_j;\mathbf{h},j)$ | Total number of competing reproductive juveniles in a patch which currently contains $j$ helpers |
| $d$ | Juvenile dispersal probability |
| $f_i$ | Fecundity of a mother breeding in a patch which currently contains $i$ helpers |
| $\phi_0,\phi_1,\phi_2$ | Parameters of the fecundity function $f_i$ |
| $h_i$ | Probability that a juvenile born from a resident mother develops as a helper in a patch currently containing $i$ helpers |
| $h_i^\bullet$ | Probability that a juvenile born from a mutant mother develops as a helper in a patch currently containing $i$ helpers |
| $h_i^\circ$ | Probability that any juvenile born on a mutant mother's patch develops as a helper in a patch currently containing $i$ helpers |
| $\mathbf{h}$ | Strategy vectors $\mathbf{h} = [h_0, h_1, \ldots h_{n_{\mathrm{h,max}}}]$ for the helping tendency expressed by offspring born from resident mothers |
| $\mathbf{h}^\bullet$,$\mathbf{h}^\circ$ | Strategy vectors for the average helping tendency expressed by offspring born from focal mutant mothers and offspring born from any mother in the mutant's patch |
| $m_i$ | Mortality probability of an adult breeder (see section S2.6) |
| $n_\mathrm{b}$ | Number of breeders in each local patch |
| $n_\mathrm{h}$ | Number of helpers in a local patch (varies among patches) |
| $n_{\mathrm{h,max}}$ | Maximum possible number of helpers in each local patch |
| $Q_{i,t}$ | Coefficient of consanguinity between two breeding mothers on a patch containing $i$ helpers at time $t$ |
| $\hat{Q}_i$ | Coefficient of consanguinity between two breeding mothers on a patch containing $i$ helpers at time $t$ |
| $r_{\mathrm{local},i}$ | Relatedness between focal mutant breeder in a patch containing $i$ helpers and any of the breeders in the focal's patch (including herself) |
| $s_{i\to j}$ | Probability that a patch which contains $i$ helpers at time $t$ will contain $j$ helpers at time $t+1$ (see eq. 1) |
| $u_i$ | Stable class frequency of patches currently containing $i$ helpers |
| $v_i$ | Reproductive value of an adult breeder in a patch currently containing $i$ helpers |
| $w_{ij}$ | Expected number of offspring born who successfully establish themselves in a patch with $j$ helpers born from a mutant focal mother in patch with $i$ helpers |

Table S1: Notation of the model on the evolution of helping.

probabilities between all the different classes of mutants

$$\mathbf{B} = \begin{bmatrix} w_{00} & w_{01} & \cdots & w_{0n_{\mathrm{h,max}}} \\ w_{10} & w_{11} & \cdots & w_{1n_{\mathrm{h,max}}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_{\mathrm{h,max}}0} & w_{n_{\mathrm{h,max}}1} & \cdots & w_{n_{\mathrm{h,max}}n_{\mathrm{h,max}}} \end{bmatrix}, \tag{S3}$$

while

$$\mathbf{A} = \mathbf{B}|_{\mathbf{h}^\bullet=\mathbf{h}^\circ=\mathbf{h}} \tag{S4}$$

denotes the resident transition matrix which comprises the transition probabilities evaluated at the population average.

## S2.2 Selection gradients

Recapping from the main text, we obtain the following term for the selection gradient $\mathcal{H}_k$

$$\mathcal{H}_k = V_k \sum_{i=0}^{n_{\mathrm{h,max}}} \sum_{j=0}^{n_{\mathrm{h,max}}} v_i u_j \left[ \frac{\partial b_{ij}}{\partial h_k^\bullet} + \frac{\partial b_{ij}}{\partial h_k^\circ} r_{\mathrm{local,j}} \right] \Bigg|_{\mathbf{h}^\bullet=\mathbf{h}^\circ=\mathbf{h}}, \tag{S5}$$

where $V_k$ is a term that is proportional to the amount of additive genetic variance in the helping tendency $h_k$. Next, $v_i$ and $u_j$ are the individual reproductive values and stable class frequencies of adult breeders which are in a patch with $i$ helpers, which are obtained from the dominant left and right eigenvectors of the resident transition matrix (see eq. [S4]). Finally, the relatedness coefficient $r_{\mathrm{local},j}$ reflects the relatedness between a focal adult breeder and all breeders in the local patch including herself (see eq. [S7]).

### S2.2.1 Selection gradient for unconditional help

The model for unconditional helping behaviour is identical to the model with developmental plasticity, with the exception that we replace $h_i$with $h$ for all $i \in 0, 1, \ldots, n_{\mathrm{h,max}}$. Consequently, the selection gradient for the unconditional strategy $\mathcal{H}$ is given by

$$\mathcal{H} = V \sum_{i=0}^{n_{\mathrm{h,max}}} \sum_{j=0}^{n_{\mathrm{h,max}}} v_i u_j \left[ \frac{\partial b_{ij}}{\partial h^\bullet} + \frac{\partial b_{ij}}{\partial h^\circ} r_{\mathrm{local,j}} \right] \Bigg|_{h^\bullet=h^\circ=h}. \tag{S6}$$

## S2.3 Relatedness

The relatedness coefficient $r_{\text{local},j}$ reflects relatedness between a focal breeder and all breeders in the local patch including herself, which is given by

$$r_{\text{local},j} = \frac{1}{n_{\text{b}}} + \frac{n_{\text{b}}-1}{n_{\text{b}}}\hat{Q}_j, \tag{S7}$$

With probability $1/n_{\text{b}}$ the same breeder is sampled, so that identity is equal to 1. Alternatively, with probability $(n_{\text{b}}-1)/n_{\text{b}}$, two distinct breeders are sampled, and the probability that both carry identical alleles is given by the equilibrium coefficient of consanguinity $\hat{Q}_j$ in demes containing $j$ breeders. The latter is calculated by solving the following recursion equation for all $j \in 0, 1, \ldots, n_{\text{h,max}}$:

$$Q_{j,t+1} = \sum_{k=0}^{n_{\text{h,max}}} \frac{u_k s_{k\to j}(h_k)}{\sum_{\ell=0}^{n_{\text{h,max}}} u_\ell s_{\ell\to j}(h_\ell)} g_k(h_k)^2 \left(\frac{1}{n_{\text{b}}} + \frac{n_{\text{b}}-1}{n_{\text{b}}} Q_{k,t}\right) \tag{S8}$$

where $g_k(h_k)$ is the probability that a philopatric juvenile successfully establishes itself as an adult breeder, or

$$g_k(h_k) = \frac{n_{\text{b}} f_k (1-d)(1-h_k)}{C(h_k; \mathbf{h}, k)}. \tag{S9}$$

## S2.4 Numerical solutions

According to a standard result (e.g., [52–54]), the trait values in the next generation are then given by

$$\begin{bmatrix} h_{0,t+1} \\ h_{1,t+1} \\ \vdots \\ h_{n_{\text{h,max}},t+1} \end{bmatrix} = \begin{bmatrix} h_{0,t} \\ h_{1,t} \\ \vdots \\ h_{n_{\text{h,max}},t} \end{bmatrix} + \begin{bmatrix} \mathcal{H}_0 \\ \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_{n_{\text{h,max}}} \end{bmatrix}, \tag{S10}$$

where $C$ reflects genetic variation due to mutation. To find the convergence stable evolutionary strategy [35, 55], we then iterate the above dynamic from the point $\mathbf{h}_{t=0} = [0.01]$ until convergence, where $|h_{i,t+1} - h_{i,t}| < 1 \times 10^{-8}$, using an algorithm written in C++ (see main text). During each timestep $t$, we solve for the equilibrium values of the left and right eigenvectors and the coefficients of consanguinity, given the updated values of $\mathbf{h}_t$. For the evolution of unconditional helping, eq. (S10) is given by $h_{t+1} = h_t + \mathcal{H}$.

## S2.5 Individual-based simulations

We also ran some individual-based simulations to corroborate our results. To this end, we simulated a population of $n_\mathrm{p} = 2500$ patches, each containing $n_\mathrm{b} = 2$ hermaphroditic breeders. At the start of each generation, each breeder chooses a random sperm donor among the $n$ individuals in the local patch (including itself). Upon fertilization, each breeder produces a number of offspring according to the same equation as in the analytical model, $f_j = (1/n_\mathrm{b})(\phi_0 + \phi_1 j^{\phi_2})$ (see the section "Evolutionary dynamics" in the main text), where we choose to multiply $f_j$ by a fecundity parameter $K = 60$ to prevent local extinctions. Each offspring has $n_{\mathrm{h,max}} + 1$ unlinked, haploid and autosomal gene loci, which correspond to the helping tendencies $h_0, h_1, \ldots, h_{n_{\mathrm{h,max}}}$. Upon inheritance, each locus independently mutates with probability $\mu = 0.01$, which involves adding a random number from a normal distribution with mean 0 and variance 0.0004 to its current allelic value (i.e., a continuum-of-alleles-model, [56]).

The number of helping offspring $n_{\mathrm{juv,help},j}$ produced by a parent with helping trait $h_j$ is then drawn from a Poisson distribution, with mean $f_j h_j$ while the number of reproductive offspring is given by $f_j - n_{\mathrm{juv,help},j}$. A proportion $d$ of all reproductive offspring is added to a pool of dispersers, which are randomized and then evenly distributed over all the $n_\mathrm{p} = 2500$ patches, while the remainder of reproductive offspring competes for breeding positions in the natal patch.

As the fecundity $f_j$ is very large in the individual-based simulations, the number of helpers is always much larger than the number of helping positions available. To approximate the analytical model therefore, we assume that all juvenile helpers experience a mortality stage (reflecting, for example, mortality during queing), during which individuals die with a probability of $m_\mathrm{juv} = 0.95$ before becoming an adult helper. The total number of helpers in the local patch in the next timestep is then the total number of surviving helping juveniles produced by all breeders. Subsequent to helper recruitment, we fill the vacant breeding positions by randomly sampling from the philopatric and immigrant juvenile reproductives, after which the cycle is repeated. Simulations ran for 40 000 generations, which was a sufficient amount of time for values of $h_i$ to reach their equilibria. Simulations are coded in C++ and the source code is available on the first author's website.

## S2.6 Overlapping generations

Here we relax the assumption of nonoverlapping generations by assuming that adult breeders die with probability $0 < m \leq 1$. For the sake of tractability, we assume that helpers are recruited anew during each timestep. The modified version of eq. (2) is then

$$\begin{aligned} w_{ij} = {} & (1-m) s_{j\to i}\left(h_j^{\circ}\right) \\ & + f_j\left(1-h_j^{\bullet}\right)\left[\frac{n_\mathrm{b} m (1-d)\, s_{j\to i}\left(h_j^{\circ}\right)}{C\left(h_j^{\circ};\mathbf{h},j\right)} + d\sum_{k=0}^{n_{\mathrm{h,max}}} u_k \frac{n_\mathrm{b} m s_{k\to i}(h_k)}{C(h_k;\mathbf{h},k)}\right]. \end{aligned} \tag{S11}$$

The other equations are identical, except for the recursions of the coefficients of consanguinity (eq. [S12]), which are now given by

$$Q_{j,t+1} = \sum_{k=0}^{n_{\mathrm{h,max}}} \frac{u_k s_{k\to j}(h_k)}{\sum_{\ell=0}^{n_{\mathrm{h,max}}} u_\ell s_{\ell\to j}(h_\ell)} \left[ (1-m)^2 Q_{k,t} + 2(1-m) m g_k(h_k) \left( \frac{1}{n_\mathrm{b}} + \frac{n_\mathrm{b}-1}{n_\mathrm{b}} Q_{k,t} \right) \right.$$
$$\left. + m^2 g_k(h_k)^2 \left( \frac{1}{n_\mathrm{b}} + \frac{n_\mathrm{b}-1}{n_\mathrm{b}} Q_{k,t} \right) \right]. \tag{S12}$$

The first part in straight brackets reflects a scenario where two breeders are sampled which have both survived (each with probability $1-m$), since the previous timestep and the probability that both have identical alleles is given by $Q_{k,t}$. The second part reflects the probability that one breeder is newborn, who has replaced a breeder that died (with probability $m$), while the other breeder has survived. In this case, the newborn breeder is born from the other sampled breeder with probability $1/n_\mathrm{b}$ (hence both sampled breeders carry identical alleles with probability 1), while with probability $(n_\mathrm{b}-1)/n_\mathrm{b}$ it was born from one of the other breeders (hence both sampled breeders carry identical alleles with probability $Q_{k,t}$). Finally, both sampled breeders are newborn with probability $m^2$, which recapitulates the expression in eq. [S12].